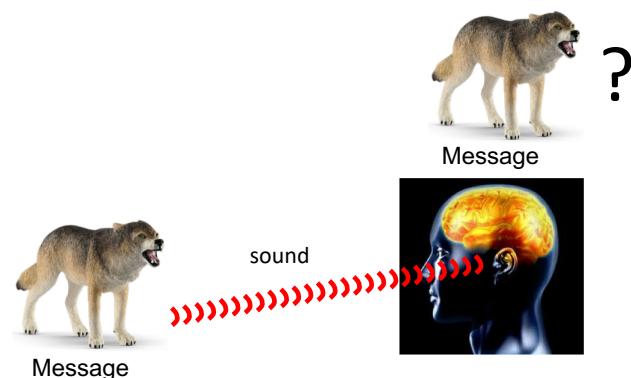
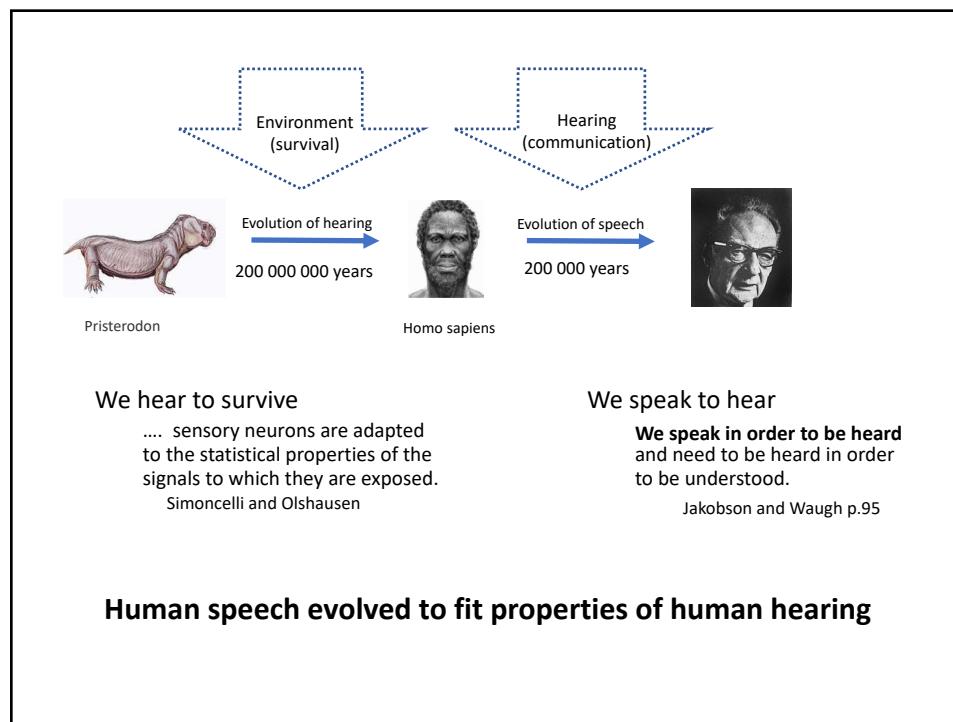
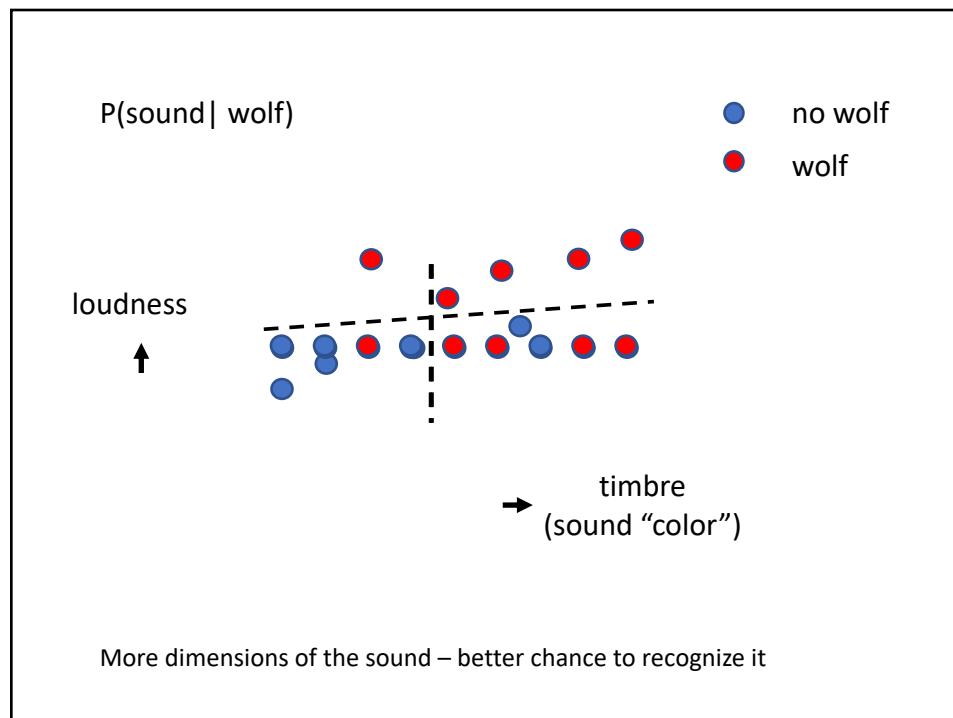


Speech

Hynek Hermansky
Electrical and Computer Engineering
Hackerman 324Fp

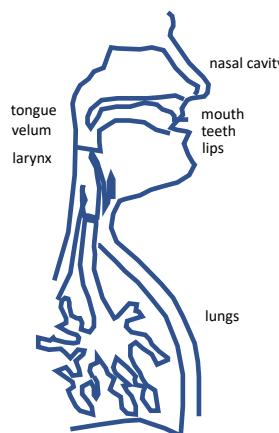


$$P(\text{wolf} | \text{sound}) \approx P(\text{sound} | \text{wolf}) \times P(\text{wolf})$$



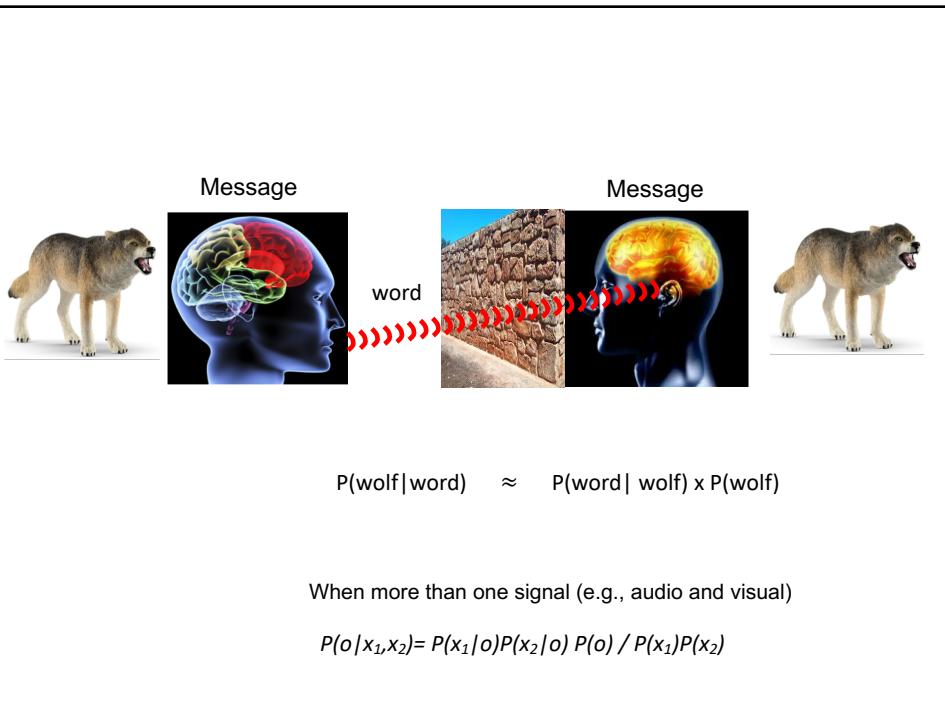
Human vocal tract

means for generation many different sounds (many dimensions)



breathing
eating
biting

speaking?

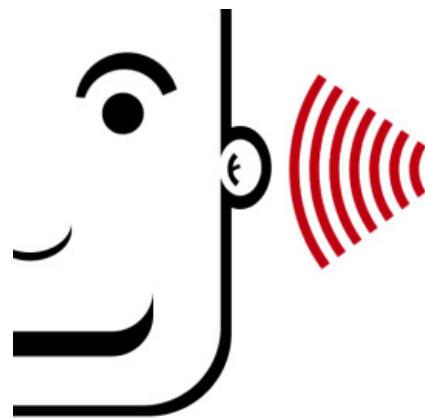


McGurk effect

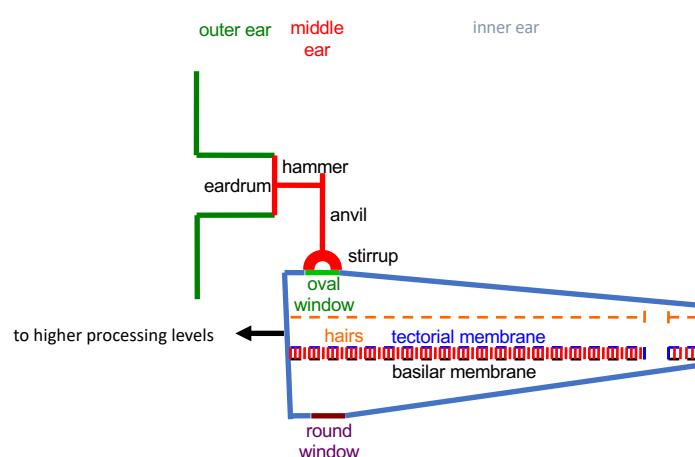
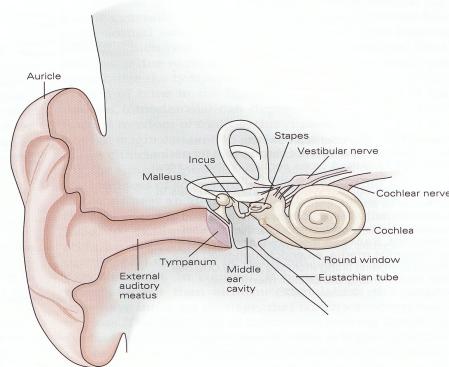
acoustic /ba/ and visual /ga/ yields /da/ or /tha/

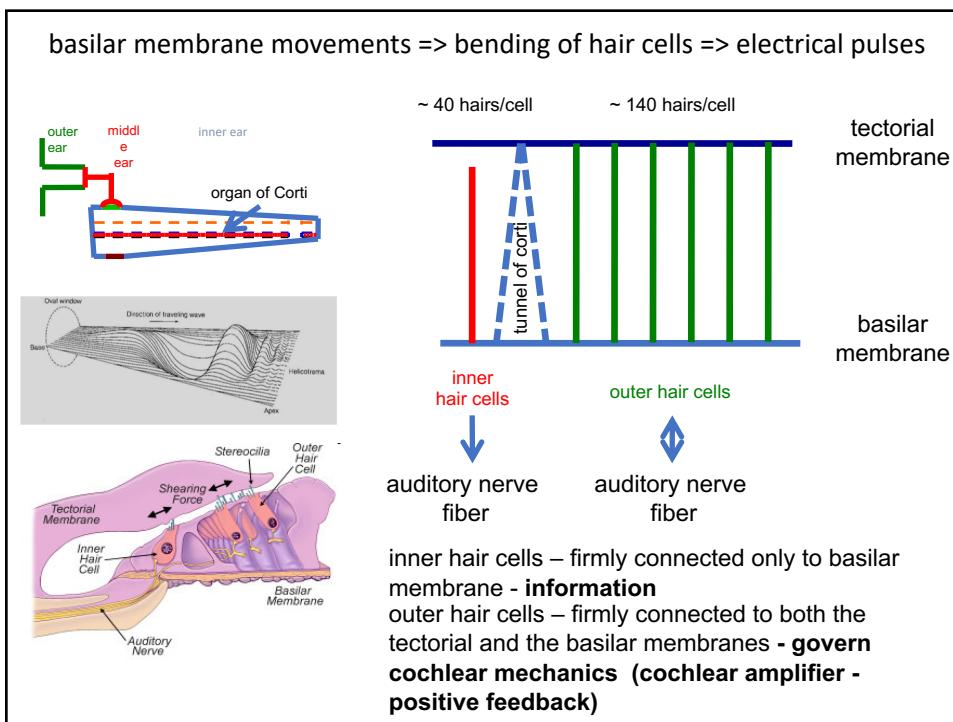
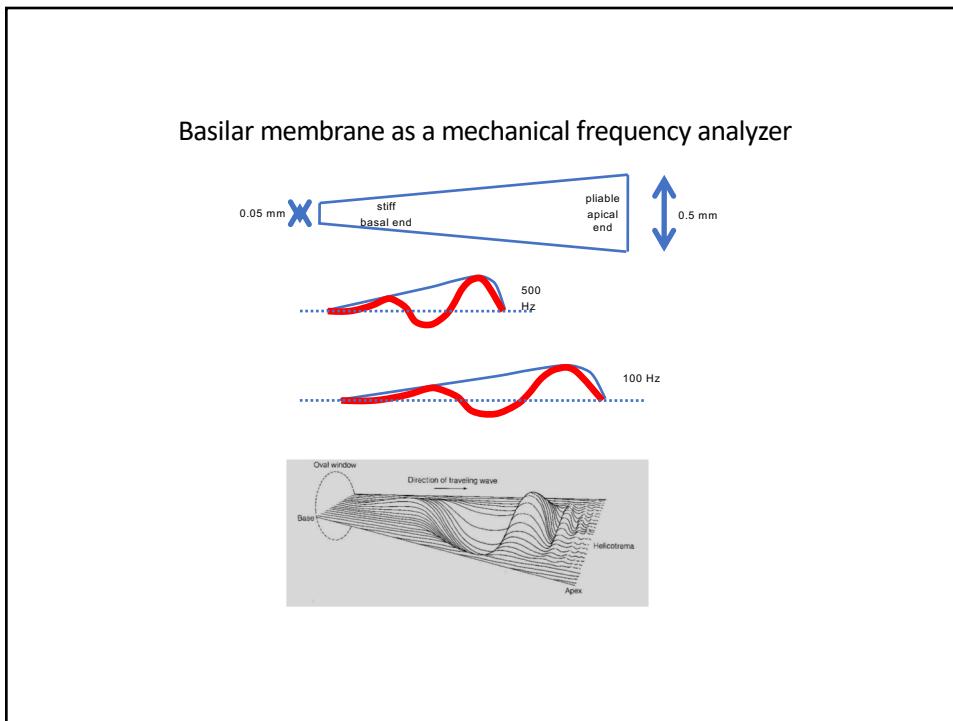


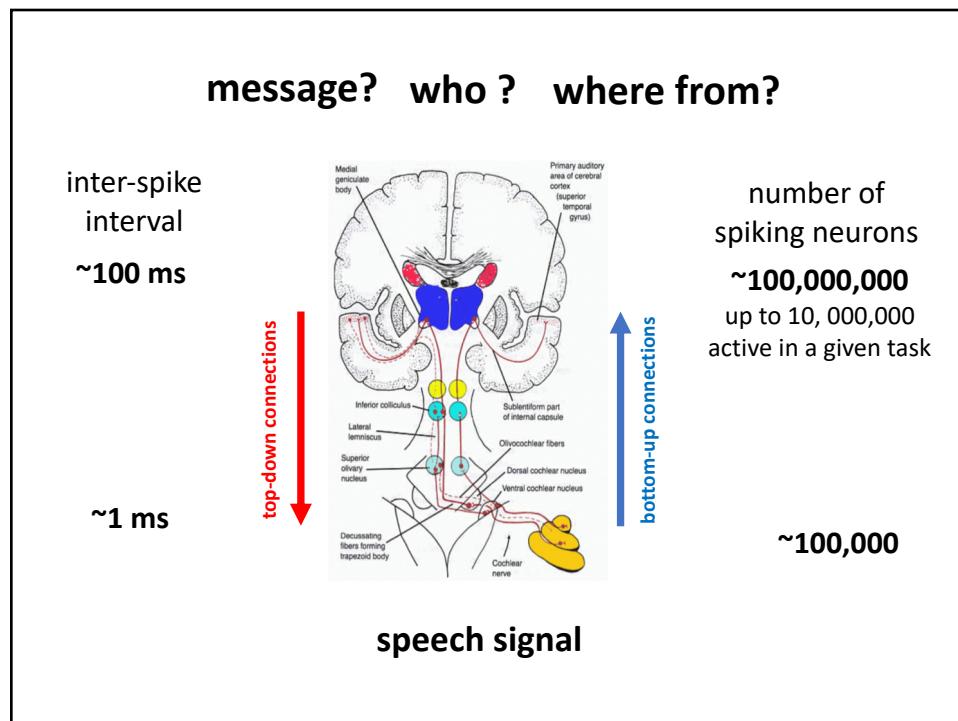
HEARING



Physiology of Hearing

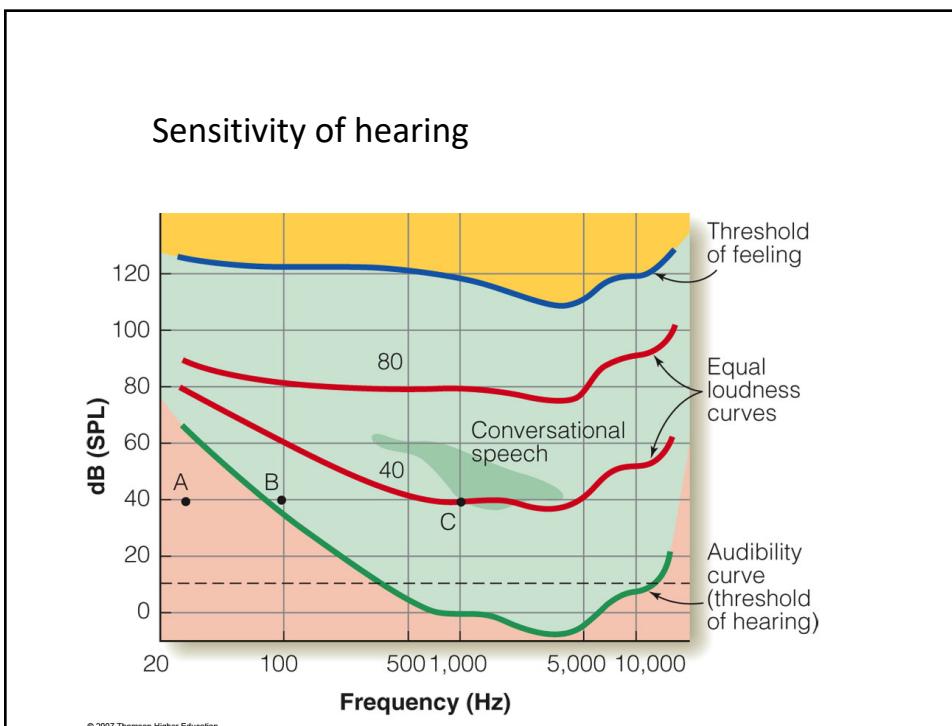
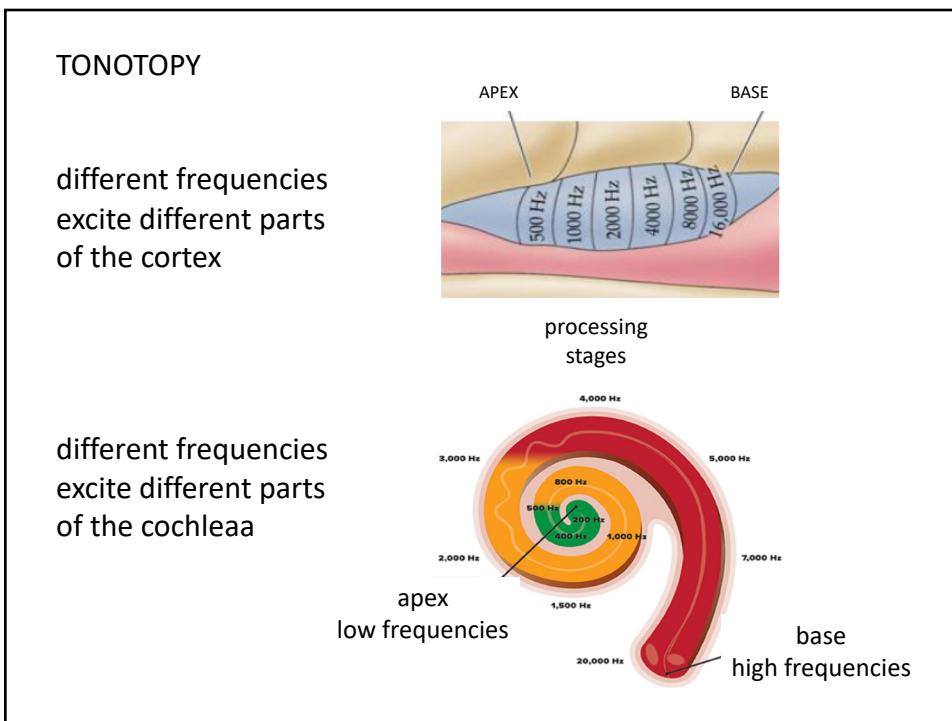




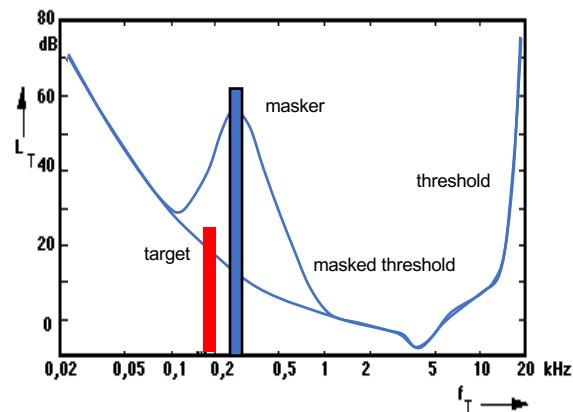


- massive increase in number of neurons from lower processing levels to cortex
- decrease in average spiking rates from periphery to cortex
- spikes in cortex are sparse (< 5% of cortical neurons active at any moment)

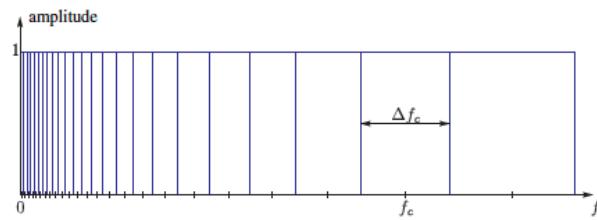
Hromadka et al PLOS Biology 2008



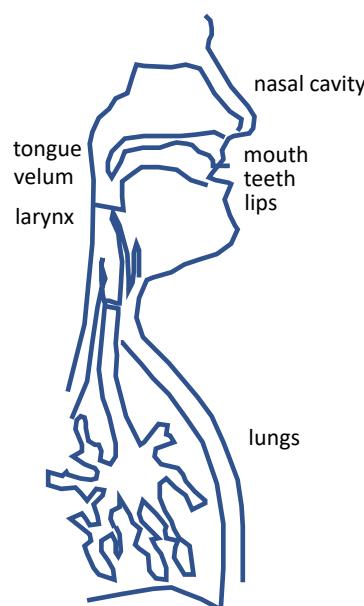
Simultaneous masking



Frequency selectivity of hearing (Critical bands of hearing)



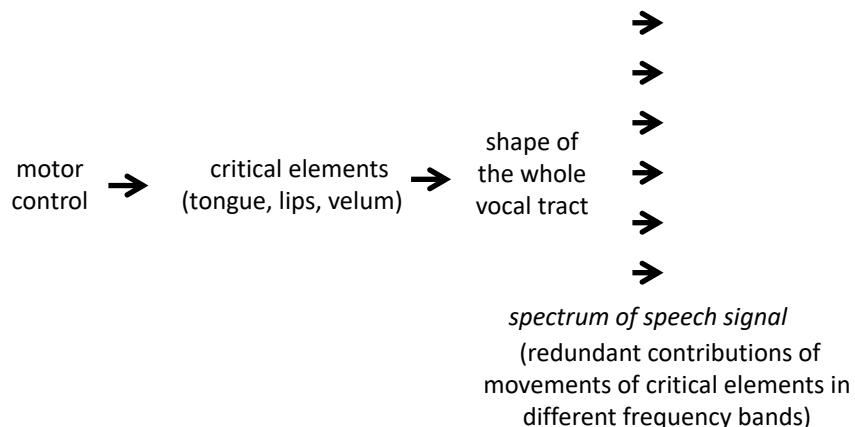
SPEAKING



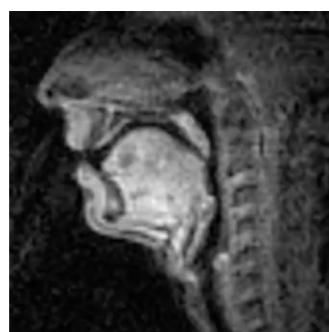
breathing
eating
biting

speaking?

INFORMATION ABOUT TRACT SHAPES DISTRIBUTED IN FREQUENCY



INFORMATION ABOUT TRACT SHAPES DISTRIBUTED IN TIME



from Sri Narajanan

movements of vocal organs are
rather sluggish

intended speech sounds

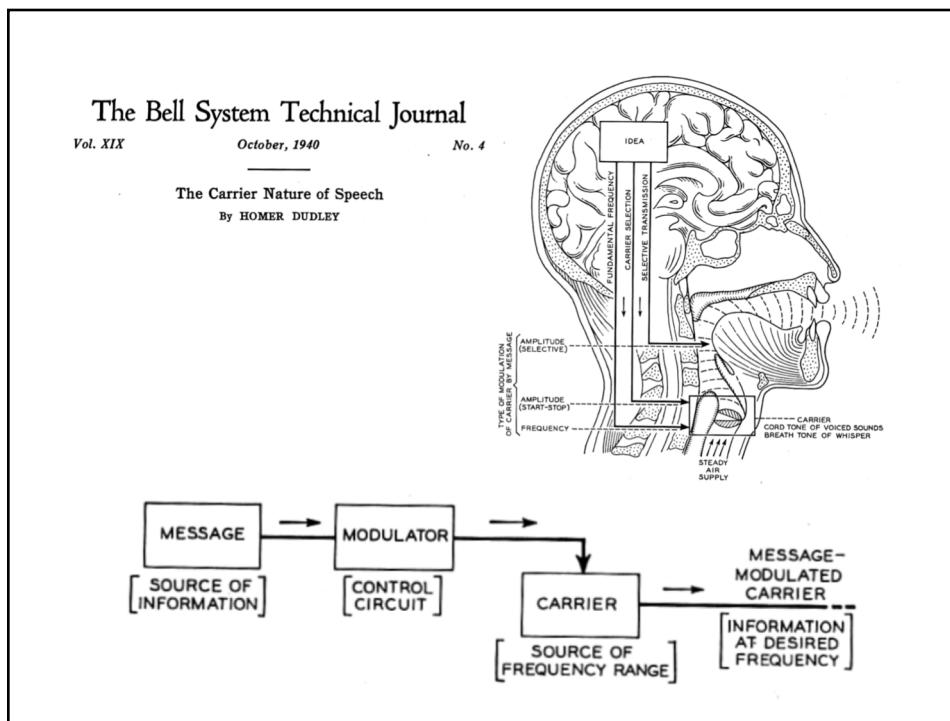


sluggishness of vocal organs



produced speech sounds





Carrier nature of speech (Dudley 1940)

```

graph TD
    A[message in] --> B["movements of  
vocal tract"]
    B --> C[modulator]
    C --> D["voiced or unvoiced  
carrier to make the tract"]
    D --> E["message  
modulated  
carrier"]
  
```

Linear model of speech production (Chiba and Kajiyama 1942)

source → filter → filtered source signal

